CLAIMS

I claim:

1. A method for managing a cache associated with a computer device in a computing environment, the method comprising the steps of:

associating each of a plurality of cached contents in the cache with generation information corresponding to a level of resources used to create that cached content; and

managing the cache based on the generation information of each content.

2. The method of claim 1, wherein the generation information for each content identifies a cost of generating that content.

3. The method of claim 2, wherein, in the associating step, the generation information for each content is associated with that content using at least one tag.

4. The method of claim 2, wherein the generation information for each content further identifies at least one of the following: an access time associated with that content, an access frequency associated with that content, and a generation time duration associated with the generation of that content.

5. The method of claim 1, wherein the managing step includes:

storing a new content in the cache if the cache is not full; and

if the cache is full, comparing generation information associated with the new content with the generation information associated with each of the plurality of cached contents; and

replacing one of the plurality of cached contents with the new content based on the results of said comparing step.

6. The method of claim 5, wherein, if the results of said comparing step identify multiple cached contents which may be replaced, the replacing step includes:

selecting one of the multiple cached contents based on a predetermined selection scheme; and

replacing the selected cached content with the new content.

7. The method of claim 6, wherein, if the results of said comparing step identify no cached content to be replaced, the method further comprises:

maintaining the cached contents in the cache.

8. The method of claim 1, wherein the computer device is an application server in the computing environment.

9. The method of claim 1, wherein the computer device is a proxy server in the computing environment.

10. The method of claim 1, wherein the plurality of cached contents represent dynamic computer pages.

11. A method for processing a content request using a cache of a servicing device in a computing environment, each of a plurality of cached contents in the cache associated with resource information corresponding to a level of resources used to create that cached content, the method comprising the steps of:

receiving by the servicing device the content request;

searching the cache of the servicing device for the requested content;

creating the requested content if the requested content is not available from the cache based on results of the searching step, the created content including resource information corresponding to a level of resources used to create that content;

attempting to cache the created content based on the resource information of the created content and the resource information of the cached contents; and

outputting by the servicing device the requested content.

12. The method of claim 11, wherein the resource information for each content identifies a cost of creating that content.

13. The method of claim 11, wherein the resource information for each content is associated with that content using at least one tag.

14. The method of claim 11, wherein the plurality of contents represent computer page information.

15. The method of claim 12, wherein the generation information for each content further identifies at least one of the following: an access time associated with that content, an access frequency associated with that content, and a generation time duration associated with the generation of that content.

16. The method of claim 11, further comprising:

receiving by a second servicing device the requested content output from the outputting step;

determining whether to cache said received content in a second cache of the second servicing device based on the resource information associated with said received content; and

returning to a user said received content according to the content request.

17. The method of claim 16, wherein the first servicing device is an application server in a communications network, and the second servicing device is a proxy server in the communications network.

18. A device for managing a cache associated with the device in a computing environment, wherein the cache includes a plurality of cached contents, each content associated with generation information corresponding to a level of resources used to create that cached content, and wherein the device manages the cache based on the generation information of each content.

19. The device of claim 18, wherein the generation information for each content is associated with that content using at least one tag.

20. The device of claim 18, wherein the generation information for each content identifies a cost of generating that content.

21. The device of claim 20, wherein the generation information for each content further identifies at least one of the following: an access time associated with that content, an access frequency associated with that content, and a generation time duration associated with the generation of that content.

22. The device of claim 20, wherein the device stores a new content in the cache if the cache is not full, and if the cache is full, the device evaluates the generation cost for each of the plurality of cached contents, detects if any cached content is associated with the generation cost which is lower than a generation cost associated with the new

content, and replaces the detected cached content with the new content based on the detection results.

23. The device of claim 22, wherein, if the detection results indicate that multiple cached contents are found, the device selects one of the multiple cached contents based on a predetermined selection scheme, and replaces the selected cached content with the new content; and wherein, if the detection results indicate that no cached content is detected, the device maintains the plurality of cached contents in the cache.

24. The device of claim 18, wherein the plurality of cached contents represent computer page information.

25. The device of claim 18, wherein the device is either an application server or a proxy server on a comminations network.

26. A system for processing a content request in a computing environment, the system comprising:

a first cache for storing a plurality of first contents, each of the first contents associated with resource information corresponding to a level of resources used to create that first content; and

a first server, associated with the first cache, for receiving the content request, searching the first cache for the requested content, and creating the requested content

if the requested content is not available from the first cache, the created content including resource information corresponding to a level of resources used to create that content, wherein the first server attempts to cache the created content based on the resource information of the created content and the plurality of first contents.

27. The system of claim 26, wherein the resource information for each content identifies a cost of creating that content.

28. The system of claim 27, wherein the generation information for each content further identifies at least one of the following: an access time associated with that content, an access frequency associated with that content, and a generation time duration associated with the generation of that content.

29. The system of claim 26, wherein the resource information for each content is associated with that content using at least one tag.

30. The system of claim 26, wherein the plurality of first contents represent computer page information.

31. The system of claim 26, further comprising:

a second cache for storing a plurality of second contents, each of the second contents associated with resource information corresponding to a level of resources used to create that content; and

a second server, associated with the second cache, for receiving the created content output from the first server, determining whether to cache said received content in the second cache based on the resource information associated with said received content and the second contents, and transmitting said received content to a user.

32. The system of claim 31, wherein the first server is an application server on a communications network, and the second server is a proxy server on the communications network.

33. The system of claim 32, wherein the resource information for each content identifies a cost of creating that content.

34. The system of claim 33, wherein the generation information for each content further identifies at least one of the following: an access time associated with that content, an access frequency associated with that content, and a generation time duration associated with the generation of that content.

35. A computer readable program product embodied on computer readable media, for implementing a method of managing a cache associated with a computer device in a computing environment, the product comprising:

first computer-readable program code means for associating each of a plurality of cached contents in the cache with generation information corresponding to a level of resources used to create that cached content; and

second computer-readable program code means for managing the cache based on the generation information of each content.

36. The product of claim 35, wherein the generation information for each content identifies a cost of generating that content.

37. The product of claim 36, wherein the first computer-readable program code means associates each content with the generation information using at least one tag.

38. The product of claim 36, wherein the generation information for each content further identifies at least one of the following: an access time associated with that content, an access frequency associated with that content, and a generation time duration associated with the generation of that content.

39. The product of claim 35, wherein the second computer-readable program code means includes:

third computer-readable program code means for storing a new content in the cache if the cache is not full;

fourth computer-readable program code means for comparing generation information associated with the new content with the generation information associated with each of the plurality of cached contents, if the cache is full; and

fifth computer-readable program code means for replacing one of the plurality of cached contents with the new content based on the results of said comparison.

40. The product of claim 35, wherein the computer device is an application server in the computing environment.

41. The product of claim 35, wherein the computer device is a proxy server in the computing environment.

42. The product of claim 35, wherein the plurality of cached contents represent dynamic computer pages.